UDC: 004.93 DOI: https://doi.org/10.15407/jai2021.02.27

INTELLIGENT SYSTEM FOR ASSESSING THE HARMFULNESS OF FOOD PRODUCTS BASED ONTHE PROCESSING OF TEXTUAL AND GRAPHIC INFORMATION

[1]V. Sineglazov, [2]O. Kozak
[1]V. M. Glushkov Institute of Cybernetics of National Academy of Sciences of Ukraine
40 Glushkov ave., Kyiv, Ukraine, 03187
[1]National Aviation University, Ukraine
1 Lubomyr Husar ave., Kyiv, 03680
[2]National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute," Kyiv, Ukraine
37 Peremohy ave., Kyiv, 03056

[1]https://orcid.org/0000-0002-3297-9060
[2]https://orcid.org/0000-0002-7541-9282

Annotation. The paper substantiates the need to assess the harm of food for consumers with chronic diseases or allergies, which is important to prevent possible deterioration of the disease or eliminate acute allergic reactions of the human body to harmful ingredients present in the product. It is proved that currently there is no convenient intelligent system that could recognize the composition of products on the Ukrainian market, provide product characteristics and assess the harmfulness of the product. It is proposed to use food labels and packaging as primary sources of food information that is available to the consumer. It is shown that the printed information on the packages is presented in text-graphic form. The development of a mobile system as a software solution for the detection and analysis of textual and graphical information on the composition of products based on the use of artificial intelligence methods is proposed and substantiated. The block diagram of the intelligent mobile system for detection and analysis of food composition has been developed. The MSER algorithm is used to select text regions on the input image matrix in the presented algorithmic software. The solution to the problem of character recognition was based on the use of convolutional neural network MobileNet-V2, which is currently the best option in the classification of images by mobile applications that do not have a server part, and therefore large computing resources. Alignment of text on the image was carried out using the method of finding a rectangle with the smallest area Developed algorithms for grouping words. A decision support algorithm has been proposed to assess the harmfulness of products. The developed system allows personalized selection of food for each individual user, ie, the assessment of the composition of products is calculated taking into account the state of health of use, existing threats, diseases, restrictions or allergies.

Keywords: analyzing product composition; assessment of the harmfulness of food; decision-making algorithm; intellectual system; text detection.

ІНТЕЛЕКТУАЛЬНА СИСТЕМА ОЦІНКИ ШКІДЛИВОСТІ ХАРЧОВИХ ПРОДУКТІВ НА ОСНОВІ ОБРОБКИ ТЕКСТОВО-ГРАФІЧНОЇ ІНФОРМАЦІЇ

[1]В. М. Синєглазов, [2]О. С. Козак
[1]Інститут кібернетики імені В. М. Глушкова Національної академії наук України, Україна
пр. Академіка Глушкова, 40, м.Київ, 03187
[1]Національний авіаційний університет, Україна
пр. Любомира Гузара, 1, м.Київ, 03680
[2]Національний технічний університет України «Київський політехнічний інститут ім. Ігоря Сікорського», Україна
пр. Перемоги, 37, м.Київ, 03056

[1]https://orcid.org/0000-0002-3297-9060
[2]https://orcid.org/0000-0002-7541-9282

Анотація. У роботі обґрунтовано необхідність отримання оцінки шкоди харчових продуктів для споживачів з хронічними захворюваннями або алергією, що важливо для запобігання можливому погіршенню перебігу захворювання або усунення гострої алергічної реакції організму людини на шкідливі інгредієнти, присутні в продукті. Доведено, що наразі не існує зручної інтелектуальної системи, яка могла б розпізнавати склад продуктів, представлених на українському ринку, надавати характеристики продуктів та оцінку шкідливості продукту. Пропонується використовувати етикетки та упаковку харчових продуктів як первинні джерела інформації про харчовий продукт, яка є доступною для споживача. Показано, що друкована інформація на упаковках представлена у текстово-графічному вигляді. Запропоновано та обґрунтовано розроблення мобільної системи як програмного рішення для виявлення та аналізу текстово-графічної інформації складу продукції на основі використання методів штучного інтелекту. Розроблено структурну схему інтелектуальної мобільної системи для виявлення та аналізу складу харчових продуктів. Для виділення текстових регіонів на матриці вхідного зображення у представленому алгоритмічному забезпеченні використовується алгоритм MSER. Розв'язання задачі розпізнавання символів виконувалося на основі використання згорткової нейронної мережі MobileNet-V2, що є на сьогодні найкращим варіантом у задачах класифікації зображень мобільними додатками, які не мають серверної частини, а отже великих обчислювальних ресурсів. Вирівнювання тексту на зображенні здійснювалося використанням методу знаходження прямокутника з найменшою площею. Розроблено алгоритми групування слів. Для оцінки шкідливості продуктів запропоновано алгоритм підтримки прийняття рішень. Розроблена система дає можливість персоналізованого підбору продуктів харчування під кожного індивідуального користувача, тобто оцінка складу продуктів вираховується з урахуванням стану здоров'я користувача, наявних загроз, хвороб, обмежень або алергій.

Ключові слова: аналіз складу продукту; оцінка шкідливості їжі; алгоритм прийняття рішень; інтелектуальна система; розпізнавання тексту.

Introduction

To take care of your health and prevent health problems, it is necessary to pay special attention to food products, before buying them, and especially to the composition of these products. Food products have a complex structure and a large number of ingredients of different chemical composition. The quality of the whole food product depends on the properties of the ingredients of the food composition.

Preserving agents, stabilizers, and emulsifiers, flavoring agents are used daily and accumulate in the human body over a long period of life. For example, sweeteners such as cyclamates (E952), aspartame (E951), saccharin (E954), which are found in many foods such as cookies and biscuits, sauces, ketchup and mayonnaise, carbonated beverages, sour milk desserts, yogurts, kefir, have the permissible amount of consuming per day and are not recommended for consuming over a long period [25]. Also, the well-known palm oil, which is still used in food (margarine, some cheeses, cakes with buttercreams, chocolate, chocolate sweets, etc.) due to a large amount of saturated fatty acids, causes cardiovascular disease and is used to make production cheaper [13, 26]. Food additives and their effects on the body can be listed further and further.

Problem formulation

According to the fact that most of the population does not read the composition of products at all, and people who read it does not understand it at all, there is a need to create a system that quickly and easily will provide consumers with all information about ingredients in the food composition, its characteristics, and give an assessment of the harmfulness of the product by food packaging.

The main requirements for the system are speed, convenience and availability; the absence of manual search, but instead the availability of means of automatic processing of the composition on the matrix of the input image; the ability to calculate a personalized assessment of harm to the user, taking into account his warnings, restrictions and preferences; the ability to obtain the characteristics of each available ingredient in the composition; availability of a database for storing the entities of this subject area; have a large database of products; ability to work without a network connection.

The mobile application will use the phone's camera to obtain images of product labels. So the system must have a software module for image

processing and retrieval, conversion of textual information in the image into text, and therefore have a module for intelligent detection and recognition of text from images.

Detecting text on images of food packaging, as well as on images in living scenes in general, differs from detecting text on document images and is one of the most difficult tasks due to a large number of image defects such as quality, noise, distortion, low-contrast background, tilt, reflection, noise and stretch, font and text size in such images [18].

To extract and recognize the text from the image of the package with the composition of the food product for further analysis, it is necessary to solve a number of the following subtasks:

- it is necessary to use mathematical algorithms for extracting a set of rectangles representing text regions, by characteristic features;
- deleted regions that contain textual information must be grouped into larger regions of words and lines;
- each region candidate of the symbol from the set of detected regions, which is represented by a pixel matrix, must be processed separately so the matrix representation of the symbol image must be intellectually recognized to the text version of the symbol;
- characters that have been recognized must be combined into words;
- the words obtained from the previous paragraph must be compared with the words of the dictionary and fill in the missing or unrecognized letters, or correct mistakes in words, after which the output is the text of the list of food ingredients obtained after recognition, which will be further processed by the acceptance algorithm solutions to assess the harmfulness of the product.

Analysis of recent research

Existing text detection algorithms are divided into two groups: region-based methods and connected components-based [1, 10]. Region-based methods analyze texture and identify areas of text based on it [1], while methods based on connected components extract candidate symbols based on edge detection or cluster color analysis. The processing of images by region-based methods causes a loss of speed because the image must be processed on several scales, especially if all calculations are performed on a mobile device [21].

Analyzing approaches to detecting symbols on food labels was found that such image objects have the same features for their detection, there is no need to use region-based methods that will re-process the image and require high computational costs for speed. Therefore, to solve the problem of removing text from the image of product labels, a group of methods based on connected components was chosen [1].

Among the methods based on connected components, for example, the Takahashi et al method extracts component text areas using a canny edge detector, then the extracted components are analyzed using a region adjacency graph [16]. The method presented by Zhu et al. works using a nonlinear local binarization algorithm to extract connected components based on several types of exceptional component features, which include special features such as geometric features, features of contrasting edges, characteristics of arrangement of details, stroke features statistics, patterns of form, etc. [13].

One of the most suitable methods for detecting text on label images is the method of detecting the most stable extreme regions (MSER), presented by Matas et al [16]. The method works by analyzing the contrast of image pixels to find image regions represented as a set of pixels that can be detected with high repeatability because they have a certain difference (from non-text regions), intensity, and stability [17], and are called extreme regions; in the images of food labels - text contrasting regions.

Machine learning methods of character classification in the image, in particular, artificial neural networks, are the best solution to the problem of recognizing text regions.

Machine learning methods that recognize symbols in images include the k-Nearest Neighbor method, the Random Forest method, the Decision Tree method, and others [7]. The most well-known method of machine learning, which has proved itself well in solving problems of character recognition, is the method of support vector machine (SVM) [20].

Today, to solve image recognition problems, researchers have focused primarily on deep learning architectures, which include recurrent neural networks (RNNs), including long-term memory architecture, and, of course, convolutional neural networks (CNN). So, we decided to use convolutional neural networks MobileNetV2, due to the proven high accuracy and efficiency of image recognition in mobile applications.

The structure of an intelligent mobile system for detecting and analyzing the composition of food products

The main aim of the study is to create an intelligent system presented in the form of a mobile application for assessing the harmfulness of food composition, based on the analysis of textual and graphical information of the manufacturer; creation of a reusable software module for detecting and recognizing text on images of product packaging. The purpose of the system is to provide the user with a decision support tool to prevent existing or potential health problems.

The system is divided into five levels, which interact closely with each other. The first level – presentation level is a user interface; the business logic level is the implementation of the subject area; the infrastructure level is software for intelligent text recognition of the composition; platform level represents the interaction of platform components, such as notification, camera, geolocation with system components; data level is a database and interaction with it.

The presentation level includes three components: the ‘personal account’ component, the component for displaying scan results, and the historical data view component. The personal account component allows the user to enter personal data such as restrictions, allergies, health problems, preferences, which will be taken into account when assessing the harmfulness of a food product. So, if there is an ingredient that is part of the user's restrictions, it will have the highest level of harm, and the system will inform the user about the presence of this ingredient in the food product. The component for displaying scan results is a presentation component that displays a product hazard assessment and a list of products with their characteristics. The historical data viewer component stores and presents information about previously scanned foods for quick access and resource savings for image and composition reprocessing. The ‘personal account’ component interacts with the database integration component to store information entered by the user. The component for displaying the results of scanning interacts with the component for analyzing the composition of products and displays the result of assessing the harmfulness of the product.

The business logic level includes a component of product composition analysis and represents the algorithmic support of the system – the decision-making algorithm. Based on the text, recognized from the image, the algorithm analyzes the ingredients and calculates the assessment of the harmfulness of the food product.

The infrastructure layer is software for processing and recognizing text in images with the composition of products.

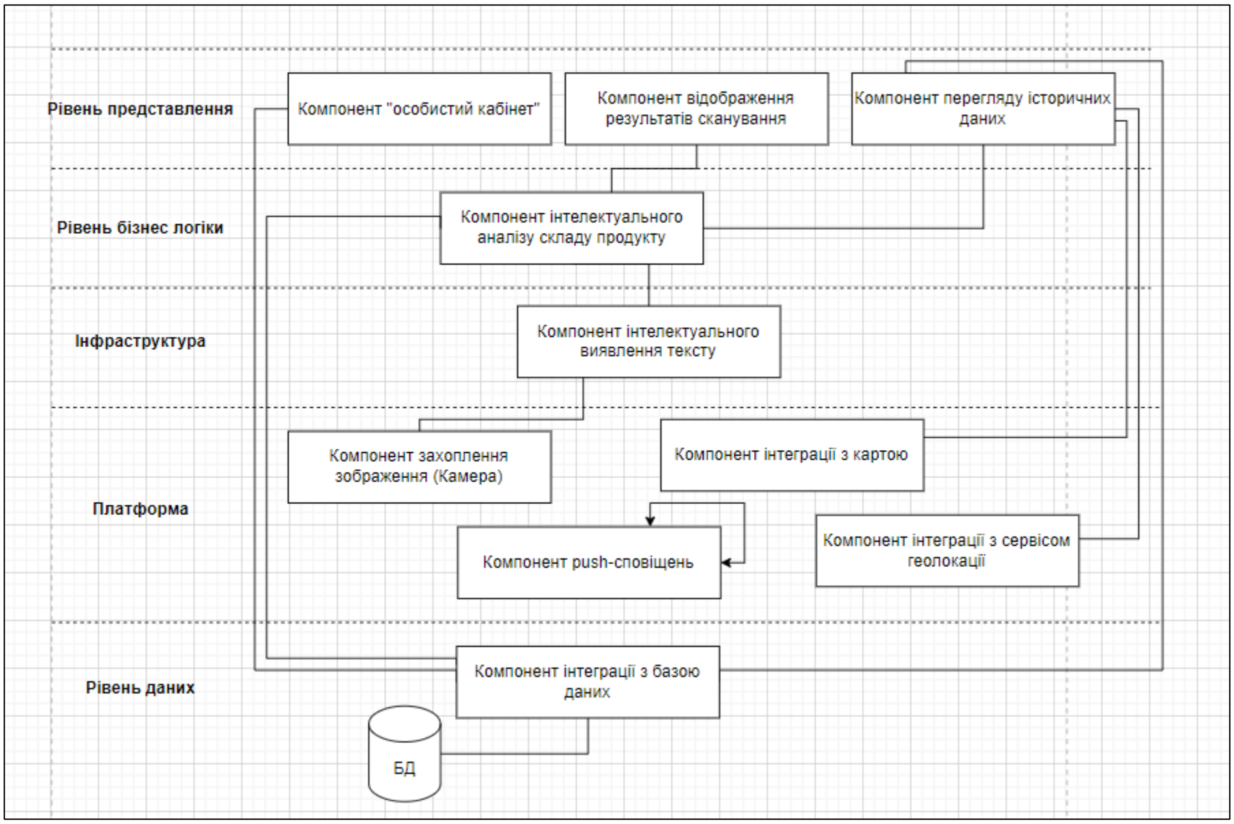

Figure 1. The structure of the intelligent system

Therefore, the image processing component performs pre-processing: contrast adjustment, image binarization to increase the speed of further recognition. The text detection component finds the letters in the image and extracts the text of the product composition in the form of an array of images. The text recognition component receives a set of single-letter images from the previously described component and recognizes letters using a neural network classifier.

The text analysis component deals with the division of the text into lines, composing letters into words. If it is necessary, the word is filled with missing letters based on the dictionary of the database.

The platform level includes the interaction of levels with platform mechanisms for ease of use of the application. The image capture component interacts with the camera and uses the camera to capture images. The push notifications component is responsible for sending reminder notifications to the user. The map integration component is used for graphical map display. The map interacts with the historical data view component and is needed to store locations of stores where food product has been scanned. The geolocation integration component makes it easy to determine the user's current location to store relevant data in the historical data view component. The data level includes a database integration component and gives other components access to data access mechanisms.

First of all, pre-processing of the input image is required: conversion of a multi-channel (color) image into a single-channel (black and white) and binarization of the image. Binarization by the process of converting the image matrix into binary using threshold approaches. Each pixel of the image is classified as white for the background and black for the text. Binarization makes it easier to detect signs of symbolic regions.

The development of the presented intelligent system for detecting and assessing the harmfulness of the composition is based on the

developed algorithmic software, which has the following structure:

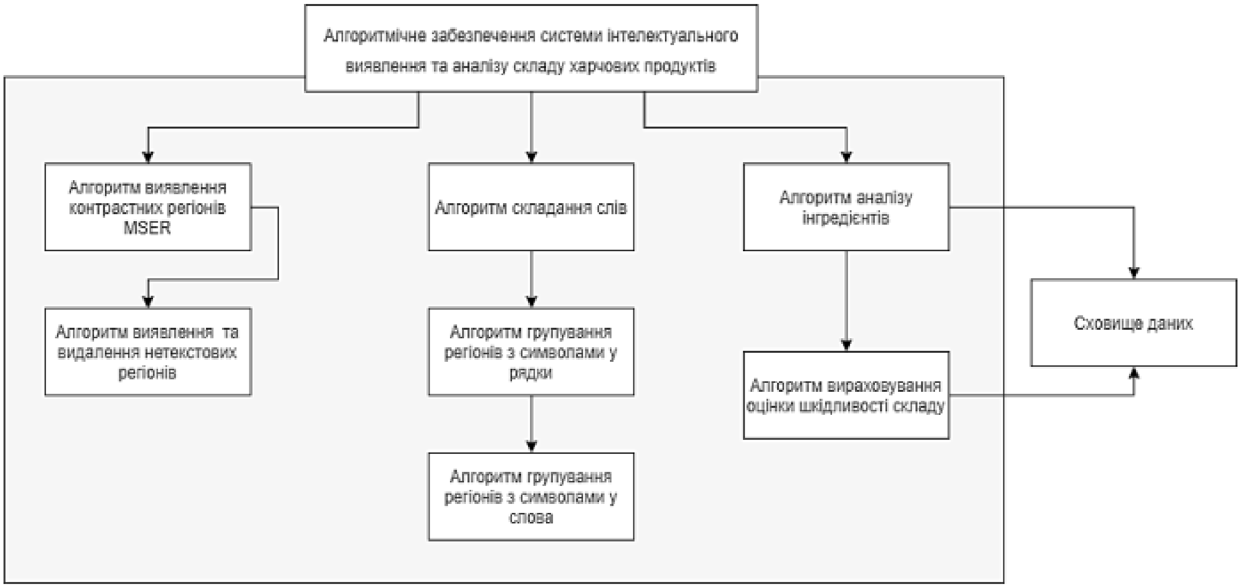

Figure 2. The structure of algorithmic software

Regionextraction

The MSER algorithm is used in the presented algorithmic software to select text regions on the matrix of the input image. The detector finds stable areas among the extreme areas of the image [2]. The extreme region is a linked region that corresponds to a certain gray level threshold, and the gray levels of all pixels in that region are greater than the threshold, while the gray levels outside this region are less than the threshold.

For an image with a given intensity, we assume that its gray level range is equal to [0,1]; n equal-interval thresholds of the gray level are set as
$\{\eta_i | \eta_{i+1} = \eta_i + \Delta, \eta_i \in [0,1], i = 1,2 \ldots, n\}$,
where Δ (delta) means the interval, and η_i is the i-th threshold.

The extreme region Q is the region of the image where for all $p \in Q, q \in \partial Q : I(p) > I(q)$ (maximum intensity region) or $I(p) < I(q)$ (minimum intensity region), whereI is a collection of sets of picture's levels,p is a certain point of the region Q, q is a certain point of the boundary of the region ∂Q [17].

That is, all pixel values in the regions are either strictly darker or strictly brighter than the values at the boundary where the intensity threshold is η.

$$\begin{cases} S\left(Q_{\eta_i}\right) \in [a,b] \\ q_{\eta_i} = \frac{|S(Q_{\eta_{i+1}})-(Q_{\eta_i})|}{(Q_{\eta_i})} < \varepsilon \end{cases},$$

where $S\left(Q_{\eta_i}\right)$represents the square of the region;

[a, b] – square size range;

q_{η_i} – the degree of change in the square of the extreme region Q_{η_i};

$\varepsilon$is the upper limit of the degree of square change [18].

The most stable region can be mathematically represented as:

$$Q_* = \arg\min_{Q_{\eta_i}} \{q_{\eta_i}, i = 1,2, \ldots\},$$

that is, from the set $\{Q_{\eta_i}, i = 1,2, \ldots\}$, the value $Q_{\eta_i}$is chosen, which has the corresponding q_{η_i}, which is the smallest in the set [18].

The output is a set of rectangles that contain the most stable extreme areas, ie candidate symbols.

The non-text areas detected by the MSER algorithm that needs to be filtered can be divided into the following groups: defective contrast areas that have a large height to width ratio, or width to the height ratio; regions with geometric shapes that are not characteristic of symbolic information; regions outside the main area with text; contrasting regions created by the interior of other regions; areas that have similar restrictive frameworks (duplicates).

Heuristic filtering of non-text regions

Let the candidate regions detected by the MSER algorithm be denoted as $C = \{c_1, c_2, \cdots, c_m\}$, where m is the total number of regions. Region x is defined as $x \in C$. According to the features of the geometry of the form and lines of symbolic information, a set of heuristic rules is established and applied. The features of the image stroke include the ratio of the stroke width to the height of the region: $R_{sh}(x) = sw_x/h_x$; the ratio of the stroke width to the width of the region: $R_{sw}(x) = sw_x/w_x$, where $sw_x$is the width of the stroke widthfor the region x. The stroke width is the value of the width of the lines that make up the symbol. Character areas have small variations in stroke width, while non-text areas have larger variations. The following principle approach is used to find the width of the stroke width:

1. First, the thinning method proposed by Lam and et al [22] is used to demonstrate the skeleton of a symbol one pixel wide.

2. Then one uses the method of obtaining the feature points of the symbol skeleton by calculating the eight values of the pixels of the neighborhood of the points of the pixel of the skeleton [8] to obtain the endpoints.

3. Then, starting from the endpoint, points are formed on the skeleton using step e, which is the distance between the points starting fromPe: P1(x1, y1), P2(x2, y2),…, $P_n(x_n, y_n)$. The angle of the point P1(x1, y1)is the angle θ between the vectors $P_1P_e = (x_e - x_1, y_e - y_1)$ та $P_1P_2 = (x_2 - x_1, y_2 - y_1)$.

The angle is calculated by the following formula [15]:

$$\theta = \arccos\left(\frac{\overrightarrow{P_1P_e} * \overrightarrow{P_1P_2}}{\overrightarrow{|P_1P_e|} * \overrightarrow{|P_1P_2|}}\right)$$

4. The stroke width, which is determined by the length of the segment a_i, is drawn through the point $P_i$perpendicular to $\overrightarrow{P_iP_{i+1}}$and bounded by the contour of the stroke. The length of the segment $a_i$is the width $w_i$of the stroke on the intervalP_iP_{i+1}.

After calculating the stroke width, set the limits:$\frac{1}{6} < R_{sh}(x) < \frac{1}{9}$, $0.3 < R_{sw}(x) < 3$.

The next step is to use a heuristic filter according to the geometric features of the symbol region. If the aspect ratio is defined as $R_{hw}(x) = h_x/w_x$, then to filter non-text regions we set the value of the ratio as $0.1 < R_{hw}(x) < 10$. The ratio between the diameter of the region, which is defined as $d = \sqrt{w^2 + h^2}$and the average value of the strike width is defined as $R_{sd}(x) = sw_x/d_x$, is set within $R_{sd}(x) < 10$.

Algorithm for removing duplicates and areas created by the interior of other regions using intersection metrics over the merge

These algorithms for filtering defective regions and regions with geometric shapes that are not characteristic of symbolic information are effective for these tasks. But the problem of duplicate regions leaves. So was decided to create an algorithm for removing duplicates based on IoU metrics. Interception over Union is a metric used to obtain an index of similarity between two arbitrary forms of objects based on the properties of the objects being compared [6].

The IoU metric is a way to solve the problem of removing duplicates. Due to the properties of the IoU similarity index, it makes it possible to compare the two extracted contrasting regions for repeatability.

Therefore, suppose that there are two rectangles on the coordinate plane with coordinates x_1, y_1, and x_2, y_2, respectively. Then the coordinates of the intersection are defined as: $x_{Ia} = \max(x_1, x_2)$; $y_{Ia} = \max(y_1, y_2)$;$x_{Ib} = \min(x_1, x_2)$;$y_{Ib} = \min(y_1, y_2)$.

The intersection area S_I is calculated by the formula for the product of the difference of the coordinates of the intersection: $S_I = (x_{Ib} - x_{Ia}) * (y_{Ib} - y_{Ia})$. Next we find the square for each of the rectanglesS_a and S_b as $S = (x_2 - x_1) * (y_2 - y_1)$. The formula of the union square S_U has the following form$S_U = S_a + S_b - S_I$.The similarity index is the following formula [12]:

$IoU = \frac{S_I}{S_U}$, whereде S_I isA ∩ B, a S_U – A ∪ B.

The initial threshold t is an indicator of the desired accuracy of duplicate removal. It is compared with the original similarity index. The coefficient that is higher than the initial threshold will be considered a true instance of the letter, and a lower one will be considered a duplicate that will be eliminated.

Algorithm1: Removal of duplicates using IoU metrics

Input: the set of rectangles $C = \{c_1, c_2, \dots, c_n\}$with coordinates x, y and size w, h, the threshold of the coefficient t at which two rectangles are considered duplicates.

Result: a set of rectangles with removed duplicates.

1. Depending on the number of rectangles in the input set, calculate the parameter m, which determines the number of rows and columns into which you want to divide the plane that places the rectangles, what is worth for optimizing.
2. Divide the rectangles into groups according to the cell (rectangle formed by dividing the plane into n columns and rows).
3. For all possible pairs of rectangles within the cell, calculate the similarity coefficient IoU.
4. Denote the rectangle in each pair, where the similarity coefficient IoU exceeds a given threshold t, one of the rectangles.
5. Obtain the resulting set of rectangles by filtering the rectangles that were marked in the previous step.

We will compose all the previous algorithms for deleting non-text areas and get regions that contain only symbolic information, which is presented in the following image:

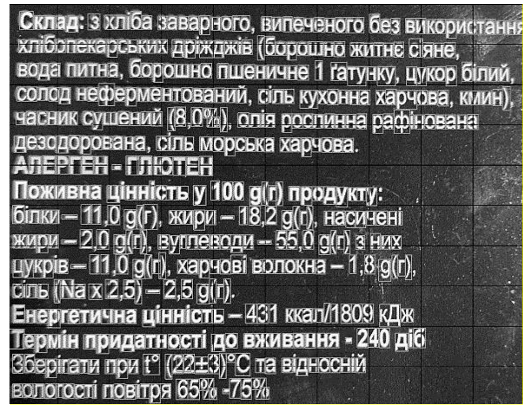

Figure 3. The result of the algorithm for deleting non-text areas

Character recognition

Detected text areas are individual image matrices that contain characters. For the problem of character recognition, the architecture of a convolutional neural network called MobileNet-V2 was chosen. This neural network is currently the best option in the classification of images by mobile applications that do not have a server part, and therefore large computing resources. MobileNet-V2 is based on the inverted structure of the residual neural network, where the residual connections are between the narrow layers (bottleneck layer).

The intermediate layer of expansion uses depthwise convolutions to filter objects as a source of nonlinearity. Depthwise layers are a replacement for standard convolution layers. Empirically, they work almost as well as conventional convolutions, and at the same time, reduce the cost less of calculations compared to traditional layers [9].

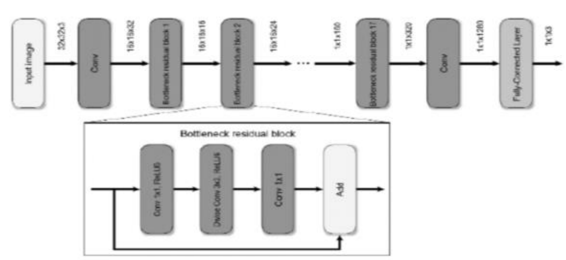

Figure 4. MobileNetV2 neural network architecture

The MobileNetV2 network architecture contains an initial convolutional layer with 32 filters, followed by 19 base blocks, called residual bottlenecks. These blocks are followed by a 1 × 1 convolution layer with an average subsampling layer. The last layer is the classification layer [14].

Composing words

With multiple rectangles that contain images of characters, you need to combine them into words.

To solve this problem, it is necessary to 1) rotate the image so that the text present in this image is aligned; 2) select rows; 3) make regions with symbols and separate regions of words.

The test skewing on the image is carried out using the method of text alignment by finding the min area rectangle [11]. The calculation of the min-area rectangle is determined by calculating the area of the bounding rectangle of the text by rotating it at different angles in a large range. The calculation takes place in two stages.During the first iteration, the rectangle rotates from α_{min} to α_{max} degrees with certain step size, $\Delta\alpha$ and the estimate of the orientation of the rectangle is calculated, in other words, the area of the rectangle at this angle is calculated. During the second iteration is the final orientation of the rectangle, calculating the area of the rectangle by rotating it in the range between the angle of the rectangle with the smallest area and the second smallest area of the adjacent rectangle.

The algorithm is as follows:

1. Suppose α_{min}, α_{max} and $\Delta\alpha$, X is a set of points forming a convex hull of a set of regions, the point of the coordinate center is the point of the center of mass of the convex hull formed by X. Set the value of $rotAngle$ as initial (α_{min}), $minArea$ is equal to the area of the current bounding box.

2. For each value of α from to α_{min} with step $\Delta\alpha$ at $\alpha \neq 0$ perform step 3 and step 4.

3. Suppose there is a line m ($y = tg(\alpha)x + b_m$) and a line n ($y = -1/tg(\alpha)x + b_n$)then the area of the limiting rectangle at the angle α is calculated by the formula:

$area = (h_1 - h_2) * (h_3 - h_4) * |cos\alpha * sin\alpha|$,where $h_1$is the maximumb_m, and $h_2$is the minimum b_n, at which the line n passes through x is X, h_3 and $h_4$are the maximum and minimum b_n, at which the line m passes through $x \in X$.

4. If$area < minArea$, then $minArea = area$, a $rotAngle = \alpha$.

5. Therefore, the angle of rotation of the region with the text is equal to $rotAngle$ [4].

After text skewing, the text in the image must be divide into lines. If one looks at the bounding boxes, one can see that their height is different. So the height of the line is mathematically determined based on information about the size of each bounding box. To find the height of the line, use the method of finding half the median of all heights of the bounding rectangles. To do this, all bounding rectangles are sorted by height, after which the median height is the middle of all heights: $H = \{h_1, h_2 \dots, h_n\}$, where n is the number of all heights, so the median value index is $m = \frac{n}{2}$. The value of the median v is equal to $h_m/2$, where m is the index of the median element.

Then, the regions that contain the image of the symbol must be grouped, taking into account a certain interval for the coordinates y (the value of half the median v). The regions are sorted by the y coordinate, after which the bounding rectangles are grouped by comparing them two by the y coordinate.

Algorithm 2:
Groupingregionswithcharactersintoline
input: setofboundingboxrectangleswithcharacters
output: setofboundingboxrectangles
begin

$H = \{\}$

forallrectangles$c \in C$
addtheheighth_c of the bounding box to H
endfor
sortlist H in ascending order;
calculate$v = h_m/2$, where$m = \frac{n}{2}$;
sort$C$ by y coordinate in ascending order;

$L = \{\}$

$l_{current} = \{\}$

foreachforallrectangles$c \in C$
if$\left|c_{prev} - c_{current}\right| \leq v$
addtothecurrentlistoflinerectangles$l_{current}$
else
clear$l_{current}$
add current$c_{current}$to$l_{current}$
end if
end foreach
for eachcoordinateofthelist$l_{current} \in L$
$y_{min} = \min(l_{current})$;
$x_{min} = \min(l_{current})$;
$y_{max} = max(l_{current})$;
$x_{max} = \max(l_{current})$;
rect $= (x_{min}, y_{min,} x_{max}, y_{max})$
add$rect$ to list $R = \{...\}$;
end for end

Combining regions into words is similar to Algorithm 2. The difference is that the value of the median quarter v is calculated as $w_m/4$, and the regions are sorted by x.

As a result, we obtain the coordinates and spatial dimensions of the bounding rectangles, which contain grouped symbols in words (Fig. 3).

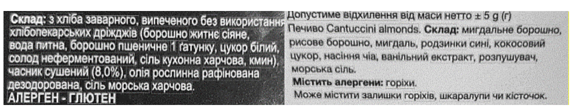

Figure 5. The result of the words grouping algorithm

Algorithm for analysis and evaluation of food hazards by its composition

The best solution for calculating the product's hazard assessment was to create an algorithm that could take into account all comments on each of the ingredients, as well as the user's comments on threats, restrictions, allergies, and intolerances, and issue an individual product hazard assessment for a specific user.

First of all, it is necessary to obtain information about each ingredient from the database. Let the set of ingredients be represented as $\mathrm{C} = \{c_1, c_2, \dots, c_n\}$, then each ingredient, if composite, ie those that include other ingredients that also have their level of threat, is represented as a graph, where the constituent ingredients of the representation

as $c_l = \{c_{l_1}, c_{l_2}, \dots, c_{l_n}, c_l \subset C\}$.

Therefore, if the ingredient is composite, it is necessary to bypass the tree of elements that make up the current ingredient and recursively subtract all threats to the tree of ingredients. These steps are necessary to avoid duplication of information at the database level, and therefore only threats to the composite ingredient that were created during its creation from child ingredients remain. Then we have a set of ingredients M, which are part of the current product $M = \{m_1, m_2, \dots, m_n\}$, where m_i is an ingredient with its characteristics (description and level of threats).

The next task is to check for the presence in the set of ingredients M of ingredients that are threatening to the user. To do this, you need to get from the database user-entered restrictions $R_c = \{r_1, r_2, \dots, r_n\}$, where R_c is a list of restrictions, allergies, user restrictions on food and ingredients.

Threatening ingredients are compared with those present in the composition and stored. Thus $R = \{r_1, r_2, \dots, r_n\}$, will contain all the ingredients from the composition, which coincide with the restrictions of the user $R_c = \{r_1, r_2, \dots, r_n\}$. At the same time, we compare the ingredients from the list of user preferences and the ingredients of the composition $P_c = \{p_1, p, \dots, p_n\}$. The current ingredient m_i, which is equal to p_j, is denoted as one that is part of the user's preferences and is stored as$P = \{p_1, p, \dots, p_n\}$.

If the current ingredient m_i is equal to r_j, then the threat level for the user $y = 0$ is set to the highest level of threat $y = 6$, the ingredient is marked as harmful to the user.

At the output we have $R = \{r_1, r_2, \dots, r_n\}$ or $R = \emptyset$; $P = \{p_1, p, \dots, p_n\}$or $P = \emptyset$; updated or unchanged list of ingredients $M = \{m_1, m_2, \dots, m_n\}$ (Fig. 4).

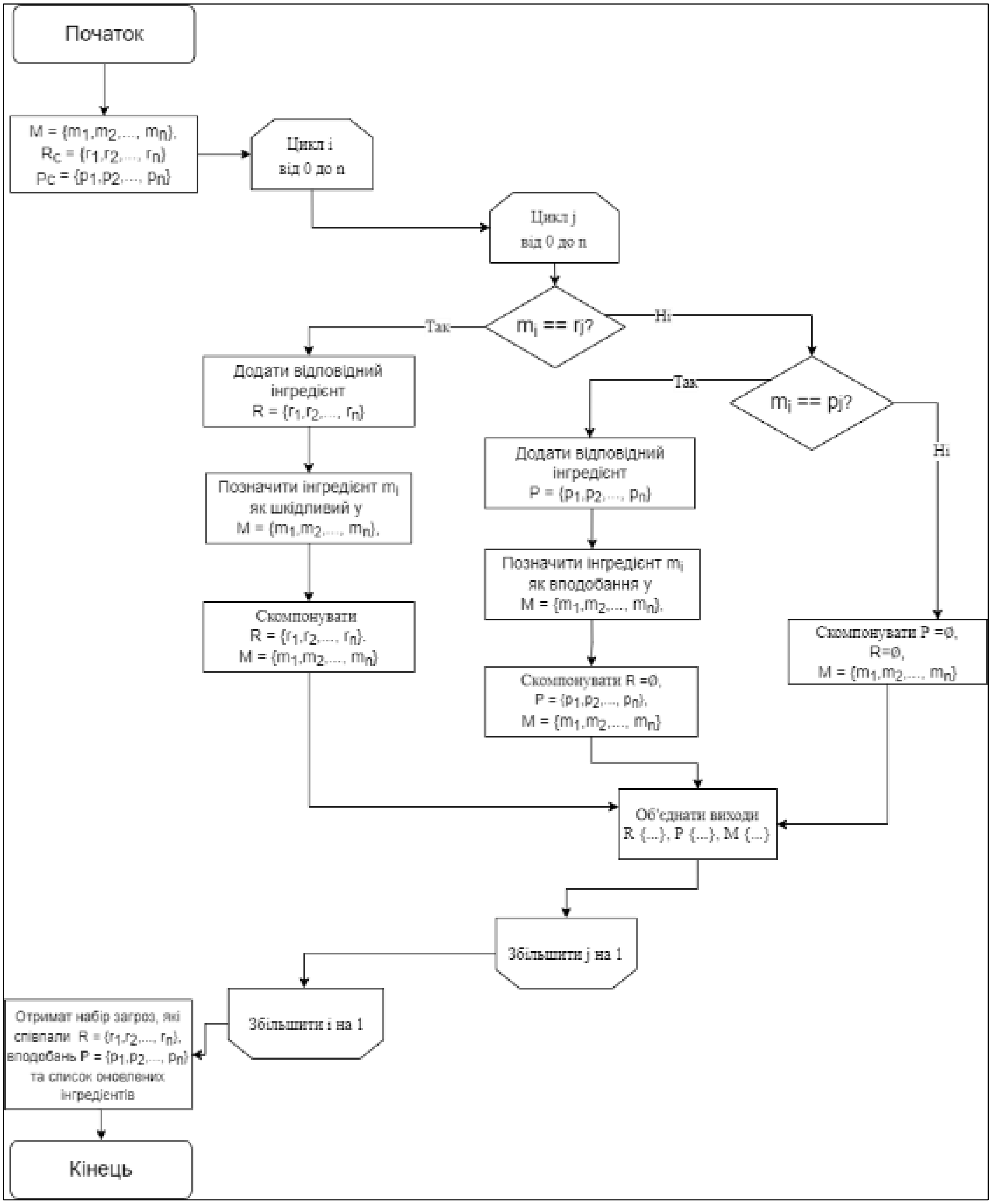

Figure 4. Block diagram of checking the presence of threatening to the user ingredients or preferences in the composition

The last step is to go through all levels of threats and calculate the hazard assessment. To assess the harmfulness, we chose the formula of the arithmetic mean weighted, for all real numbers x_1, x_2, …, x_n with weights w_1, w_2 …, w_n and is defined as:

$$x = \frac{x_1 w_1 + x_2 w_2 + \cdots + x_n w_n}{w_1 + w_2 + \cdots + w_n}$$

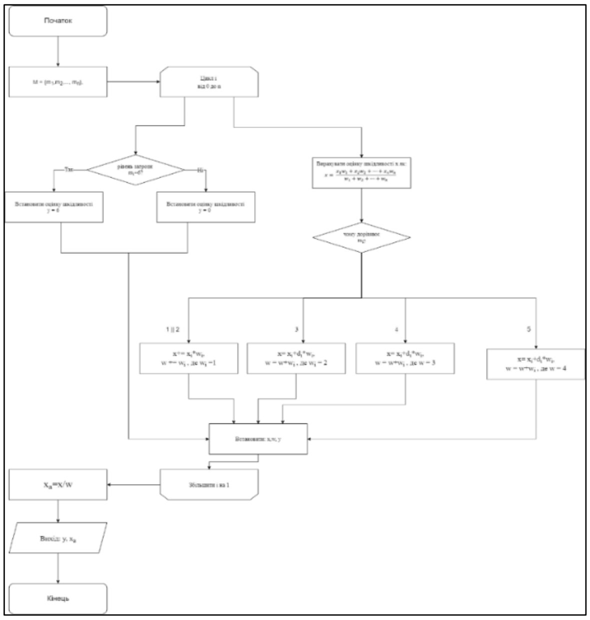

Figure 5. Block diagram of the calculation of hazard assessment

It was taken by weights: w = 1 for for level 1, 2; w = 2 for level 3, w = 3 for level 4, w = 4 for level 5.

At the output, we obtain an estimate of the harmfulness of the product (from one to six, which is converted into an alphabet, where 1 = A, 5 = F) based on the analyzed ingredients. If the maximum level of threat is 6, then the composition receives the status "the product does not fit the user's preferences" (Fig. 5).

Conclusions

The intelligent system for determining and assessing the harmfulness of food today is of great social importance, which determines its creation. An intelligent presentation system in the form of a mobile application has been developed, which is due to the requirement for system convenience and speed of use.

Therefore, the best solution was to use a mobile phone camera to scan food labels, after which the textual information about the composition of the resulting image is intellectually processed. Due to the limited computational costs that a mobile application can provide, a text recognition software module based on the use of algorithmic software and convolutional neural networks, has been created. Algorithmic software includes methods for identifying regions with text, aligning text in the image, grouping the resulting regions into word regions.

The developed system allows personalized selection of food for each user, ie, the assessment of the composition of products is calculated taking into account the state of health of use, existing threats, diseases, restrictions, or allergies. Therefore, the real-time decision support algorithm calculates the harm assessment for each user based on the product composition information.

References

1. A.HybridApproachtoDetectandLocalizeTextsinNaturalSceneImages. Yi-Feng Pan; Xinwen Hou; Cheng-Lin Liu. IEEE Transactions on Image Processing vol: 20, no: 3, pp. 800 – 813 March 2011. doi: 10.1109/TIP.2010.2070803
2. Beyond MSER: MaximallyStableRegionsusingTreeofShapes. PetraBosilj, EwaKijak, Sébastien LefèvreSeptember 2015.doi:10.5244/C.29.169. [Online] – Available: https://www.researchgate.net/publication/281565129_Beyond_MSER_Maximally_Stable_Regions_using_Tree_of_Shapes
3. Chyshkidlyvapalmovaoliya. Отримано з https://moz.gov.ua/article/health/chi-shkidliva-palmova-olija
4. Classificationoffusedfaceimagesusingmultilayerperceptronneuralnetwork, DebotoshBhattacharjee, MrinalKantiBhowmik, MitaNasipuri, DipakKumarBasu, MahantapasKundu. Отримано з https://arxiv.org/abs/1007.0633
5. Documentskewdetectionusingminimum-areaboundingrectangle.RezaSafabakhsh. Shahram Khadivi. February 2000. doi:10.1109/ITCC.2000.844226. [Online] – Available: https://www.researchgate.net/publication/384829

6_Document_skew_detection_using_minimum-area_bounding_rectangle
6. Foodlabelreading: Readbeforeyoueat. Goyal R, Deshmukh N. Отримано з https://www.researchgate.net/publication/324761439_Food_label_reading_Read_before_you_eat
7. GeneralizedIntersectionoverUnion: A Metricand A LossforBoundingBoxRegression.February 2019. HamidRezatofighi, NathanTsoi, JunYoungGwak,AmirSadeghian. Отримано з https://www.researchgate.net/publication/331371000_Generalized_Intersection_over_Union_A_Metric_and_A_Loss_for_Bounding_Box_Regression
8. HandwrittenOpticalCharacterRecognition (OCR): AComprehensiveSystematicLiteratureReview (SLR) JamshedMemon, MairaSami, andRizwanAhmedKhan. July 2020IEEE Access. doi:10.1109/ACCESS.2020.3012542. [Online] – Available:https://www.researchgate.net/publication/343273822_Handwritten_Optical_Character_Recognition_OCR_A_Comprehensive_Systematic_Literature_Review_SLR
9. Identificationofforkpointsontheskeletonsofhandwrittenchinesecharacters. K. Liu, Y. S. Huang, and C. Y. Suen, IEEE Trans. PatternAnal. Mach. Intell., vol. 21, no. 10, pp. 1095–1100, Oct. 1999, DOI: 10.1109/34.799914
10. Informacionnayazapiska INFOSAN No. 3/2006 - Pishchevyeallergii. Vsemirnayaorganizaciyazdravoohraneniya.Отриманоз https://www.who.int/foodsafety/fs_management/No_03_allergy_June06_ru.pdf
11. Metodidentyfikaciyiharchovyxdobavok (pidsolodzhuvachiv) zmetoyuvyyavlennyafalsyfikaciyiprodukciyi / P. G. Stolyarchuk [tain.] // VisnykNacz. tehn. un-tu "HPI": zb. nauk. pr. Temat. vyp.: Novi rishennya v suchasnyhtehnologiyah. – Harkiv: NTU "XPI", 2010. – # 46. – S. 3-7.
12. MobileNetV2: InvertedResidualsandLinearBottlenecksMarkSandlerAndrewHowardMenglongZhuAndreyZhmoginovLiang-ChiehChen. Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, Liang-Chieh Chen. doi: 10.1109/CVPR.2018.00474. Отримано з https://ieeexplore.ieee.org/document/8578572
13. Natural Scene Text Detection with Multi-channel Connected Component Segmentation. Xiaobing Wang; Yonghong Song; Yuanlin Zhang. 12th International Conference on Document Analysis and Recognition, 2013. doi: 10.1109/ICDAR.2013.278
14. OntheMinimum-AreaRectangularandSquareAnnulusProblem. SangWonBae. Отримано з https://arxiv.org/abs/1904.06832
15. Optimizingintersection-overunionindeepneuralnetworksforimagesegmentation. Md Atiqur Rahman and Yang Wang. Отримано з http://cs.umanitoba.ca/~ywang/papers/isvc16.pdf
16. Pediatric Age Palm Oil Consumption.LorenzaDiGenova, LauraCerquiglini, LauraPenta, AnnaBiscarini, andSusannaEsposito. Отримано з https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5923693/
17. Real-Time and Accurate Drone Detection in a Video with a Static Background.UlzhalgasSeidaliyeva, DarynAkhmetov, LyazzatIlipbayeva, Eric T. Matson. doi:10.3390/s20143856. Отримано з https://www.researchgate.net/publication/342856036_Real-Time_and_Accurate_Drone_Detection_in_a_Video_with_a_Static_Background
18. RecognitionofHandwrittenChineseCharactersBasedonConceptLearning. Liang Xu, YuxiWang,XiuxiLi,Ming Pan. July 2019.IEEE Access PP(99):1-1. doi:10.1109/ACCESS.2019.2930799
19. Region graph based text extraction from outdoor images. Hiroki Takahashi. August 2005. doi:10.1109/ICITA.2005.235 [Online] – Available:https://www.researchgate.net/publication/4162652_Region_graph_based_text_extraction_from_outdoor_images
20. Robustwide-baselinestereofrommaximallystableextremalregions.J. Matas, O. Chum, M. Urban, and T. Pajdla. Отримано з https://cmp.felk.cvut.cz/~matas/papers/matas-bmvc02.pdf
21. ShipDetectionfromOcean SAR ImageBasedonLocalContrastVarianceWeightedInformationEntropy. WeiboHuo,YulinHuang,JifangPei,QianZhang,QinGuandJianyuYang. Отримано з https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5948720/
22. SurveyonTextDetection, SegmentationandRecognitionfrom a NaturalSceneImages. UmaKaranje, RahulDagade. InternationalJournalofComputerApplications 108(13):39-43.December 2014. doi:10.5120/18974-0472. [Online] – Available:https://www.researchgate.net/publication/287689667_Survey_on_Text_Detection_Segmentation_and_Recognition_from_a_Natural_Scene_Images
23. TextDetectionandRecognitioninImagesandVideos. DatongChen, Jean-MarcOdobez, HerveBourlard. January 2004.

Отримано з https://www.researchgate.net/publication/37433214_Text_Detection_and_Recognition_in_Images and_Videos
24. Text detection and recognition using enhanced MSER detection and a novel OCR technique. Md. Rabiul Islam; Chayan Mondal; Md. Kawsar Azam; Abu Syed Md. Jannatul Islam. 5th International Conference on Informatics, Electronics and Vision (ICIEV), 2016. doi: 10.1109/ICIEV.2016.7760054.
25. Thinningmethodologies – a comprehensivesurvey. L. Lam, S. W. Lee, and C. Y. Suen, IEEE Trans. PatternAnal. Mach. Intell., vol. 14, no. 9, pp. 869–885, Sep. 1992, doi: 10.1109/34.161346
26. Using Adaboost to Detect and Segment Characters from Natural Scenes. K. H. Zhu, F. H. Qi, R. J. Jiang, L. Xu, M. Kimachi, Y. Wu, and T.Aizawa. Отримано з http://www.imlab.jp/cbdar2005/proceedings/papers/O2-3.pdf

Received 30.11.2021
Accepted 10.12.2021